簡帛文字近音假借現象的發生原因和證明方法

2018/12/16 顧國林 （2020/03/12 修改）

[提要] 統計 6455 對簡帛假借字，發現兩個假借規則：一、近音假借發生率與音節字數呈負相關；二、同音假借發生率與音節字數呈正相關。由此可證明，非同音假借現象，是古人尋找同音聲符困難造成的，造字時已經不同音。方言影響說、時代變遷說都不能成立。

一、疑問

這幾十年，出土了很多簡帛文字，有大量假借和異文，這些假借字對，從中古音來看，大部分不同音。這種不同音現象，是什麼原因造成的呢？主要有以下幾種認識：

（1）方言。如 A 方言區造的字拿到 B 方言區使用。

（2）時代。用字時代和造字時代遠隔，語音變化造成字音參差（這個觀點雖然不符合語言學「一變全變」演化規律，但在文字學裡有依據，文後擴展部分會談及）。

（3）《廣韻》記音不可靠，漏掉了古音，漏掉了本音。

（4）假借對語音要求不高，第一次假借發生時已不同音。

下文的結論是：（4）是正確的。

例證法有不可靠之處，下面採用數學方法提供較可信的證明。

二、證明方法

通過分析造字、用字心理，設計一種探測方法：檢測近音假借發生率、同音假借發生率對音節字數的敏感程度。

我們站在古人用字的角度思考：如果一個音節有多達幾十個常用字，那麼沒必要捨近求遠去借一個近音字來用，但如果一個音節只有它一個字，那麼別無選擇，只能借近音字。同音字數量對假借的音準，有決定性影響。假設存在近音假借，規律一定是這樣的：音節的同音字越多，近音假借發生率越低，同音假借發生率越高。

用以上規律去檢驗存世假借字，如果符合，可以證實近音假借，如果不符合，可以證偽。

出土簡帛為研究假借行為提供了優秀條件，使用材料為白於藍《簡帛古書通假字大系》，這部書收集出土簡帛假借字 9655 對，是該領域集大成者，它的特點如下：

（1）全部是出土材料，年代比較集中（戰國到秦漢），前後跨度約 500 年。

（2）收錄較齊全，覆蓋了 2017 年出版前刊布的絕大部分簡帛材料。

用電腦初步統計結果是：在該書 9655 對假借字裏，前後字（借出字和借入字）都見於《廣韻》的有 6455 對，中古同音的有 1872 對，中古不同音的有 4583 對[1]。簡帛假借字有大量一對多、多對一的情況，比如用「不」字寫「丕、背、否」、用「門、昏、聞」字來寫「問」等。所以，以上的 6455 對假借裏，如果去除重複字，那麼前字是 3133 個，後字是 3413 個。下文統計遵守 6455 對的數量，同時記字次。

[1] 對於異讀的處理方法是：比較每個字的記音，如果能找到一次同音，則算同音。

三、取得結論

由於上古音節的精確面貌不知，退而求其次，使用《廣韻》音節。我們知道，小韻是同一個反切下同音字的集合，廣韻一共有 3867 個小韻，大體上可視為是中古音的音節[2]。本節把 6455 個被假借字（如女-汝中的汝）放到每個音節中作統計。

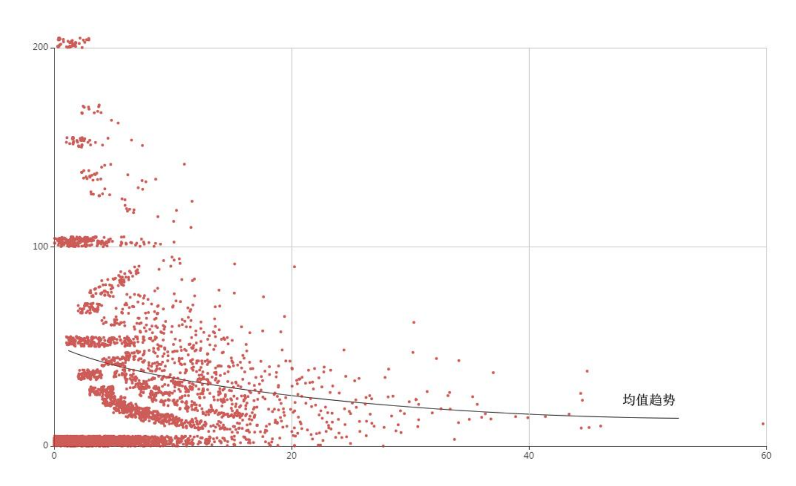


圖 A：簡帛近音假借發生率（豎軸）對廣韻小韻字數（橫軸）的下降式分佈[3]

（每個點即一個廣韻小韻，共計 3867 個）

約定 4 個概念：

（1）音節字數：指《廣韻》小韻內同音字數量，有 1-87 個不等，以 5-20 個為常見。

（2）總轄字：同字數音節轄字總和，比如有 10 個字音節有 130 個，總轄字：10*130=1300 個。

（3）總假借數：同字數音節的簡帛被假借字總和，比如 10 字音節有 130 個，總假借數是 451 個，其中中古同音者 165 個、中古不同音者 286 個。

（4）假借發生率：總假借次數/總轄字，比如 10 個字的音節，總字數是 1300 個，總假借次數是 451 個，假借發生率為 451/1300=34.7%，由於記次數，可大於 100%。還可細分為同音假借發生率、近音假借發生率。

[2] 小韻包含《廣韻》增修的重出小韻，含有少量重複音節，一般位於該韻最後，但數量很少，在統計效應中可以忽略。

[3] 在字數很少時，縱軸數值只有固定幾個，會造成大量點重疊在一起，為了更好地展示每個點，圖中對縱軸數據進行了加[0,10]的隨機偏移，對橫軸數據進行了加[0,1]的隨機偏移，這樣就能把重疊在一起的點在小範圍內擴散開，便於觀看。該圖週邊還有極少量數據點，因佔用版面太大沒有列出。

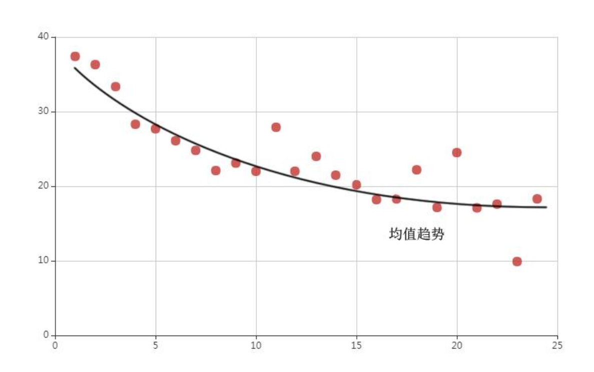


圖 1：簡帛近音假借發生率對音節字數的下降式分佈

（1）統計結論一：**近音假借發生率與音節字數呈負相關**

[現象] 圖 1（即圖 A 的統計形式）橫軸為音節字數，縱軸為近音假借發生率，分佈點取總轄字大於 200 字的音節[4]（詳細數據見文後附錄表 1）。從圖中可以看出，近音假借發生率的趨勢隨著音節字數的增加而下降，從起始的約 35%下降至 18%，下降幅度約一半。兩者的皮爾森相關性[5]數值為-0.856，顯示為「極強負相關」，表明兩者有非常明顯的負比例關係。

[解釋] 近音假借發生率隨著音節字數的增加而下降，反映的歷史面貌是：古人使用記音字，在同音字數越多的音節裏，越容易找到同音聲符，採用近音字記音的概率越低。

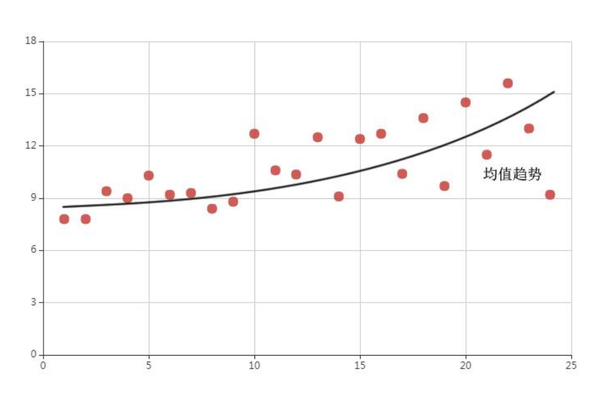


圖 2：簡帛同音假借發生率對音節字數的上升式分佈

（2）統計結論二：**同音假借發生率與音節字數呈正相關**

[現象] 圖 2 橫軸為音節字數，縱軸為近音假借發生率，分佈點取總轄字大於 200 字的音節（詳細數據見文後附錄表 2）。從圖中可以看出，同音假借發生率的趨勢隨著音節字數的增加而上升，從起始的約 8%上升至 13%，上升幅度約一半。兩者的皮爾森相關性計算數值為+0.654，顯示為「強正相關」，表明兩者有明顯的正比例關係。

[解釋] 同音假借發生率隨著音節字數的增加而上升，反映的歷史面貌是：古人使用記音字，在同音字數越多的音節裏，越容易找到同音聲符，採用同音字記音的概率越高。

[4] 樣本數量少影響精確度，統計價值低，所以小樣本常在統計中舍去。

[5] 對於數據處理，判斷相關性（第一組增大時，對應第二組是否也相應增大或相應減少）一般採用皮爾森相關性係數，文章也採用這個方法，皮爾森相關性計算的結果值範圍是[-1,1]，正數表示 A 列數據增大時 B 列也增大，負數則反之，絕對值是 1 時，表示所有數據點在一條直線上，是理想的線性關係，絕對值為 0 時，表示數據雜亂無規律，可視為沒有相關性，常用的數值參考如下：0.8-1.0（極強相關）、0.6-0.8（強相關）、0.4-0.6（中等相關）、0.2-0.4（相關）、0.0-0.2（弱相或無相關），更精確的應用還需考慮標本數量。

四、總結

以上兩圖，證明大量假借在簡帛時代第一次產生時，就已經不同音。

現象一和現象二是互補的，可以相互解釋，它們反映，古人假借用字時，同音字和近音字是並用的，而且會優先使用同音字，體現在「同音多的音節裏同音假借發生率」，在沒有辦法的情況下會使用近音字，體現在「同音字少的音節裏近音假借發生率高」。

從中古音角度看，這 6455 對假借字，同音只有 1872 對，占 29%，近音有 4583 對，占 71%，雖不能將此比例直接套用到上古，但這個比例大致能反映當時的社會情況，在上古找一個合適的同音記音字並非容易。中古的音節將近 4000 個，常用字如果按現代標準 3500 個計算的話，平均每個音節只有約 1 個常用字，上古的音節更多，每個音節擁有的常用字更少，這使得古人不得不用近音字表音，實為無奈之舉。今天在普通話裏能輕而易舉找到同音字，已是漢語音節大簡並的結果了。

最後解釋一下用《廣韻》小韻作音節單位，為什麼能在探索上古簡帛假借字中奏效。《廣韻》小韻是中古音的同音字組，繼承自上古音的同音字組，雖然演化中會有兼併，但規模上仍具有繼承性。打個比方，班級每個學生的小學成績和中學成績，對個體來說是有變動的，但從整體來看，繼承性仍然是明顯的。從統計角度，用《廣韻》小韻上推上古音節，雖然精度有損失，但結論仍然有效。

附 1：《廣韻》記音問題

一種普遍觀點認為：中古不同音的字，上古肯定不同音，假借是否近音，觀察中古音即可。

對語音史來說，這話沒有錯，因為語音史有「音變無例外」的理論支撐，但對文字史來說，這話就不準確了，因為：

（1）語言沒變，但字音變了。我們今天從《廣韻》裏查到的音，並不對應它們在上古造字時的音。比如，它是蛇初文，午是杵初文，凡是盤初文，且是祖（一說俎）的初文，在最早造字和假借時，它、午、凡、且用的是蛇、杵、盤、祖的「本音」，而不是《廣韻》裏的記音，後者是漢字假借後形成的「新音」。從上古到中古，漢字經歷了反復假借，很多本音被驅逐，《廣韻》記音已不是造字時代的原貌。

（2）諸侯國造字。漢字可能在不同方言區裡製造，有些假借在誕生方言裏同音，但傳到別的方言變得不同音。舉個現實的例子：通語裏借戴（咍韻）寫帶（泰韻），如「帶帽子（老寫法）」寫成「戴帽子（新寫法）」，通語不分咍、泰韻，屬於同音假借。後來，這種寫法傳到分咍、泰韻的吳語裏，戴就產生了異讀（老戴[te]用在姓氏和地名裏，新戴[ta]用於穿戴義）。

所以，「中古不同音的字，上古肯定不同音」在文字史裡是不準確的，假借時的真實面貌，是很難從中古音判斷的，除非把所有漢字的來龍去脈都研究清楚，目前文字學還做不到這一點，這使得對假借語音本質的探索要尋找新的方法。

參考文獻：

《宋本廣韻》 江蘇教育出版社 2008 年

白於藍 2017 年 《簡帛古書假借字大系》 福建人民出版社

李方桂 1980 年 《上古音研究》 商務印書館

鄭張尚芳 2008 年 《上古音系》 上海教育出版社

周法高等 1975 年 《金文詁林》 香港中文大學

張世超等 1996 年 《金文形義通解》 中文出版社

附 2：簡帛假借發生率和音節字數關係的數據表

音節字數	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
總轄字	537	1018	1242	1440	1565	1614	1785	1592	1476	1300	1298	1284	1066	854	1035
總近音假借數	201	370	413	408	434	421	443	352	341	286	362	282	256	183	209
近音假借發生率	37.4	36.3	33.3	28.3	27.7	26.1	24.8	22.1	23.1	22	27.9	22	24	21.4	20.2
音節字數	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
總轄字	848	595	450	361	440	357	352	322	240	200	182	189	112	174	180
總近音假借數	154	109	100	62	108	61	62	32	44	35	39	16	22	33	47
近音假借發生率	18.2	18.3	22.2	17.2	24.5	17.1	17.6	9.9	18.3	17.5	21.4	8.5	19.6	19	26.1
音節字數	31	32	33	34	35	36	37	38	40	42	44	45	47	60	87
總轄字	93	64	132	102	70	108	74	38	40	42	88	180	47	60	87
總近音假借數	16	9	31	13	22	23	9	5	4	5	15	30	3	5	4
近音假借發生率	17.2	14.1	23.5	12.7	31.4	21.3	12.2	13.2	10	11.9	17	16.7	6.4	8.3	4.6

表 1：簡帛近音假借發生率對音節字數的分佈數據

音節字數	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
總轄字	537	1018	1242	1440	1565	1614	1785	1592	1476	1300	1298	1284	1066	854	1035
總同音假借數	42	79	117	130	161	148	166	133	130	165	138	133	133	78	128
同音假借發生率	7.8	7.8	9.4	9	10.3	9.2	9.3	8.4	8.8	12.7	10.6	10.4	12.5	9.1	12.4
音節字數	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
總轄字	848	595	450	361	440	357	352	322	240	200	182	189	112	174	180
總同音假借數	108	62	61	35	64	41	55	42	22	27	17	16	19	26	32
同音假借發生率	12.7	10.4	13.6	9.7	14.5	11.5	15.6	13	9.2	13.5	9.3	8.5	17	14.9	17.8
音節字數	31	32	33	34	35	36	37	38	40	42	44	45	47	60	87
總轄字	93	64	132	102	70	108	74	38	40	42	88	180	47	60	87
總同音假借數	27	8	23	15	10	7	16	10	1	3	10	41	6	4	12
同音假借發生率	29	12.5	17.4	14.7	14.3	6.5	21.6	26.3	2.5	7.1	11.4	22.8	12.8	6.7	13.8

表 2：簡帛同音假借發生率對音節字數的分佈數據